

Interactive Tutorial 1: SNP Database Resources

The tutorial is designed to take you through the steps necessary to access SNP data from the primary database resources:

1. Entrez SNP/dbSNP
2. HapMap Genome Browser
3. NIEHS - GeneSNPs and the NIEHS SNPs Websites
4. Other Tools –PolyPhen, ECR, PolyDoms, Transfac

Note: Answers to questions from this tutorial are included at the end of this document

As a launching point, we will begin our searching at the Entrez cross-database browser. This can be accessed on the NCBI home page (<http://www.ncbi.nlm.nih.gov/>). For these exercises we will be accessing data for the gene: chemokine-like factor (HUGO name: NOS2A).

For a cross-database search:

1. Enter the gene symbol (NOS2A) into the empty box next to the ‘Search All Databases’, type NOS2A into the empty box and click on the GO button, or simply hit the return key on your keyboard.
Which NCBI database gives the most number of results
What is the database? Hint: Mouse over on the ‘?’ next to the icon and click for a popup explanation of this database.
2. On the left column note the results returned for the ‘SNP’ and ‘Gene’ database.
3. How many results were returned for the ‘SNP’ and ‘Gene’ database?
4. Why did the ‘Gene’ database return more than one result?

Entrez Gene

5. From the cross database search, click on the ‘Gene’ database icon.
6. Click on the result that corresponds to the ‘homo sapiens’ NOS2A gene.
7. NOS2A maps to which chromosome?
8. What are the genes 5’ and 3’ of NOS2A? (Hint: look at the genomic context).
9. On the far right of the page next to the NOS2A gene name and description, note the word ‘Links’(see Figure 1 below).
10. Scroll down this list and select ‘SNP: Geneview’

NCBI Entrez Gene

Search Gene for nos2a Go Clear current records only

Limits Preview/Index History Clipboard Details

Display Full Report Show 20 Send to

All: 1 Genes Genomes: 1 SNP GeneView: 1

1: NOS2A nitric oxide synthase 2A (inducible, hepatocytes) [Homo sapiens]
 GeneID: 4843 Primary source: [HGNC:7873](#) updated 21-Dec-2005

Summary

Official Symbol: NOS2A and **Name:** nitric oxide synthase 2A (inducible, hepatocytes) provided by [HUGO Gene Nomenclature Committee](#)

See related: [HPRD:01225](#), [MIM:163730](#)

Gene type: protein coding

Gene name: NOS2A

Gene description: nitric oxide synthase 2A (inducible, hepatocytes)

RefSeq status: Reviewed

Organism: [Homo sapiens](#)

Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo

Gene aliases: NOS; INOS; NOS2; HEP-NOS

Summary: Nitric oxide is a reactive free radical which acts as a biologic mediator in several processes, including neurotransmission and antimicrobial and antitumoral activities. This gene encodes a nitric oxide synthase which is expressed in liver and is inducible by a combination of lipopolysaccharide and certain cytokines. Three related pseudogenes are located within the Smith-Magenis syndrome region on chromosome 17. Alternative splicing of this gene results in two transcript variants encoding different isoforms.

Genomic regions, transcripts, and products

(minus strand) [RefSeq below](#)

NC_000017

23151662 5' 23167920 3'

NL_008625 NP_080616 isoform 1
 NL_153292 NP_080624 isoform 2

— coding region — untranslated region

Entrez Gene Home
 Table Of Contents
 Summary
 Genomic regions, transcripts...
 Genomic context
 Bibliography
 HIV-1 protein interactions
 Interactions
 General gene information
 General protein information
 Reference Sequences
 Related Sequences
 Additional Links
 Links
 Conserved Domains
 Genome
 GEO Profiles
 HomoloGene
 Map Viewer
 Nucleotide
 OMIM
 Full text in PMC
 Probe
 Protein
 PubMed
 PubMed (GeneRIF)
 SNP
 SNP: Genotype
 SNP: GeneView
 Taxonomy
 UniSTS
 AceView
 CCDS
 Ensembl
 Evidence Viewer
 GDB
 HGNC
 HPRD
 KEGG
 ModelMaker

dbSNP

1. The initial dbSNP Geneview only shows SNPs that are located in the coding region of the gene (cSNPs).
2. How many cSNPs are found in dbSNP for NOS2A? How many are validated? Under the 'Gene Model' heading, use the button selectors to view all SNPs in the 'gene region' (select that button) and then select the 'view rs' button.
3. After selecting this, the page will update and show all SNPs in this gene.
4. How many SNPs are found in dbSNP for NOS2A?
5. Note: this number will appear just above the SNP map picture of the gene.
6. How many SNPs shown have been validated by the HapMap project? i.e. have an 'H' symbol in the validation column?
7. How many SNPs have frequency data (i.e. a heterozygosity value) associated with them? Hint: count the number without this data and subtract from total
8. Click on the rs# link for the intronic SNP rs7208775. What methods were used to validate this SNP?
9. How many submitters have recorded a discovery of this SNP?
10. Click on the ss# (ss38342908) next to the 'EGP_SNPS|NOS2A-0044457' SNP submission.
11. On this page, scroll down and find the frequency data for this SNP in each of the four populations studied by this submitter (EGP_AD, EGP_ASIAN, EGP_YORUB, EGP_CEPH). What the allele frequency of the C and G allele in each of these populations?

12. Using the 'BACK' button in your browser, return to the Entrez Gene page for NOS2A.

Entrez SNP

1. Starting from the Entrez Gene page again, use the 'Links' menu on the right side to view the linkout choices and select the 'SNP' option.
2. This will automatically query the Entrez SNP database for all SNPs in dbSNP for the NOS2A gene for species you are viewing (i.e. 'homo sapiens').
3. How many SNPs are returned?
4. Below the search box and tabbed menu choices (i.e. 'Limits', 'Preview/Index', etc), the 'Display' feature menu to show this list as a 'FASTA'. The page should automatically update when you make your selection.
5. In the 'Send To' drop down menu, select the 'Text' option. The page should update the results in plain text format. This selection can be directly copied to a file on your computer.
6. Use the 'BACK' button on your browser. Alternatively this data can be "Sent To" a 'File' directly, that is saved on your computer.
7. Select the 'Limits' tab below the main search box. In the main search box type the gene name 'NOS2A'.
8. Select the following search limits from the selections on this page:
 - a. Organism: Homo sapiens
 - b. Validation: 2hit-2allele
9. After making these selections, use the 'Go' button next to the main search box to get the result.
10. How many results are returned for validated 2hit-2allele SNPs in this gene?
11. Experiment with saving these in different formats using both the 'Send To' → 'Text' option and 'Send To' → 'File' option.
12. Go to the 'Limits' option again and select the following search items:
 - a. Organism: Homo sapiens
 - b. Has Genotype: True
13. How many results are returned for SNPs with genotype data?
14. In the 'Display' drop down menu select the 'Genotype'. Genotype data for each rs# SNP will be displayed.
15. Make sure that the checkbox in the 'Limits' tab is unchecked. Finally, to demonstrate the ability of using search term fields directly in the main search box, type the following:

NOS2A[gene] AND "EGP_SNPS"[handle]

How many total entries are in dbSNP for this gene and submitted by the NIEHS EGP project (our handle is EGP_SNPS)?

HapMap Browser:

The HapMap Genome Browser is linked directly from the main page at hapmap.org by selecting 'Browse Project Data'.

1. In the main search box enter the abbreviated gene name with a wild card 'NOS*'.
2. A list of result from many genes will be displayed – select NOS2A.
3. The browser page with tracks will be presented.
4. How many HapMap SNPs in this gene?
5. Zoom in to view 20kb using the Scroll/Zoom drop down menu
6. Note the display of frequency data for each population using the pie graphs for each SNP. Click on the first HapMap SNP in this gene (rs2297516)
7. Note the allele frequencies for each population. Select the 'retrieve genotypes' link on the far right column. Genotype data for this SNP and population will be displayed
8. Use the 'BACK' button to get to the main gene view.
9. Below the gene structure image (exons), click on the 'Tracks' setting. Under the 'Analysis' subheading select 'plugin:tagSNP picker'. Next select the 'Update Image' using the button on the right side of the page.
10. SNPs correlated with rs2297516 should appear on a track. Select the first one (rs2297516). In the 'Search' section of layout (near where you entered the gene symbol) select the 'Reports & Analysis' pull down menu selection – 'Download tagSNP data' and select 'GO'.
11. Which SNPs are captured by this tagSNP?

GeneSNPs:

<http://www.genome.utah.edu/genesnps/>

1. Select 'Cell Cycle' from the 'Gene List' drop down menu.
2. Select the 'UCSC:hg16:4' Gene Model link for the gene CCNA2.
3. What is the orientation of the gene?
4. In the 'All Submitter' drop down menu, select EGP_SNPS.
5. Scroll down the list of SNPs to the first non-synonymous cSNP. In what position of the codon is this SNP?
6. On what population was this gene re-sequenced? Click on the GS link at the right. What is the genotype of the sample P012?

NIEHS SNPs.

<http://egp.gs.washington.edu>

The variation data for the NOS2A can be accessed through the search box on the top right part of the home page, or via the 'A-Z Finished Genes Directory' link in the left hand navigation bar.

Using this last link, find NOS2A in the alphabetical listing of genes.

1. Under the 'Mapping Data' section, click on the 'cSNPs' link.
2. How many non-synonymous cSNPs were discovered in this gene? What is the position in our reference sequence of the of the first synonymous SNP?
3. What was the cDNA position of this SNP?
4. In which population was it discovered?
5. Go 'BACK'. In the 'Genotyping' category, click on the 'Visual Genotype' link. An image of all the genotyping data for this gene is displayed. Using the SNP location of the synonymous SNP, determine which individual carries this polymorphism?
6. Explore other links in the 'Mapping Data', 'Genotyping Data' and 'Predictive Data' sections. The 'Linkage Data' and 'Haplotype Data' will be covered in a subsequent talk and tutorial. Compare the data from the NIEHS SNPs Environmental Genome Project to that found in the other database resources.

Genome Variation Server:

<http://gvs.gs.washington.edu/GVS/>

This can also be accessed off the SeattleSNPs home page

Note: Check that your browser is set to NOT block pop-up windows. GVS uses pop-up windows to display the images and tables.

1. Click the Gene Name search button
2. Enter 'NOS2A'
3. A table of 'Populations' and 'Submitters' will appear along with the number of genotyped SNPs for each entry.
4. Select the button for 'HapMap-CEU'. De-select AFD-CHN.
5. Select the 'Display Genotypes' button at the bottom of the page.
6. Two windows should pop-up having a visual genotype image and a table with the genotypes for every sample at every SNP site.
7. Experiment with the other buttons for displaying results (e.g. display tagSNPs, etc).
8. Parameters can be used to customize these results.

PolyPhen:

<http://tux.embl-heidelberg.de/ramensky/>

1. Enter the amino acid sequence for BRCA1 from the fasta file provided. Include the first line starting the '>' character.
2. Enter 356 for position, Q (glutamine) and R (arginine) for AA₁ and AA₂, then "Process query."

3. The polyphen site returns its prediction, based on alignments of both polypeptide sequences to sequences in the SwissProt data base, the potential of disrupting known structural motifs (coils, active sites, disulfide bridges, phosphorylation sites, etc.), and the steric changes to the three dimensional structure. This substitution is predicted to be “probably damaging” (and is a known BRCA1 mutation).
4. Navigate to the NIEHS SNPs EGP web site, use the ‘A-Z Finished Genes Directory’ link to list the “B” genes and go to BRCA1.
5. Click on the non-synonymous cSNP analysis link in the Predictive Analysis section and examine the Polyphen and Sift predictions.
6. Which substitutions are predicted by Polyphen to be damaging? Which substitutions are predicted to be intolerant by Sift?
7. Extra assignment: Determine which sample in the PDR carries the BRCA1 mutation. Hint: see NIEHS SNPs tutorial.

Polydoms:

<http://polydoms.cchmc.org/polydoms/>

8. Enter BRCA1 in the ‘Gene/Protein Symbol’ field.
9. In the Search Results section, 13 proteins (1 unique) should be listed. Click on the ‘NonSynonymous’ link of the first Gene Description (protein NP_009225).
10. Find the Q356R substitution. What non-EGP submitters have identified this variant?
11. Use the color key under the gene schematic to find the “mutation.” What substitution is coded as a mutation by OMIM?
12. In the Pathway(s) section, click on the ATM signaling pathway to view an image of BRCA1 interacting proteins from Biocarta.
13. Go back to the NIEHS_SNPs site, find the Pathway Image Maps section on the navigation bar at the left side of the page. Click on the ‘Double Strand Break Repair’ link. Mouse over ATM and the other protein icons to find other genes in this pathway scanned for variations. Note how these two resources can be used to identify SNPs in interacting genes and pathways.
14. Return to the PolyDom site and click on the Download Polydom Annotations link under Search Results. Select All and Download table. Save to a file (snpAnnotation.xls) and open using Excel.
15. Examine the table and note how many cSNPs are reported in BRCA1?

ECR browser and Transfac:

<http://ecrbrowser.dcode.org/>

<http://bimas.dcrt.nih.gov/molbio/signal>

1. Enter BRCA1 in the search field next to the submit button, then click ‘Submit’.

2. Click on the RefSeq corresponding to chr17:38449842-38530657.
3. Note the horizontal blue lines indicating the gene structure of BRCA1 isoforms. The arrows indicate the gene is shown 3' to 5' (right to left). Use the arrow buttons on the bottom left of the page (next to the < and > buttons) to flip the orientation of the gene.
4. Zoom out 1.5 X using the green “-1.5 X” button to show the genomic context around BRCA1. What is the gene shown upstream of BRCA1? What is the orientation of the gene?
5. Note the pink colored peaks in the mouse and chicken BRCA1 orthologues: Pink indicates conserved intronic regions, yellow UTR, red intergenic regions (potential enhancers), and blue exons.
6. Click on the pink region conserved in chicken to center the browser on this region.
7. Click on the “+10 X” button to zoom in on this region.
8. Click on the “Grab ECR” button.
9. Click on the pink region in the mouse corresponding to the pink region in chicken. A conserved sequence alignment of human vs. mouse will appear in a new window.
10. Swipe the human sequence at the bottom of the page and select ‘Edit’ and ‘Copy’ from the browser menu.
11. Go to the Transfac page at the URL listed above. Transfac is a database of consensus transcription factor binding sequences (TFBS).
12. Paste the sequence into the Nucleic Acid sequence window in the page. Select the ‘Mammal’ class and click the ‘submit’ button. What is the longest consensus signal sequence of the hits listed? How many of these potential transcription factor binding sites are in this ECR? Is this TFBS also found in chicken? Hint: click on the ‘Site #’ R01162.
13. Are any SNPs located in these potential TFBS? Hint: this is a difficult question to address. What is the strand the TFBS is found on? A more direct approach may be to ask if there are any potential TFBS around a specific polymorphism.
14. Go back to BRCA1 on the NIEHS SNPs EGP page and click on the ‘SNP Context’ link. Scroll down to the sequence for the insertion-deletion polymorphism at 013578. Swipe the three lines of sequence (corresponding to upstream flanking, the insertion, and downstream flanking sequence) and ‘Copy’ and ‘Paste’ it into the Nucleic Acid sequence window of the Transfac page then click ‘submit.’ Note the potential TFBS.
15. Swipe and ‘Cut’ the middle line of sequence (corresponding to the inserted sequence) from the Nucleic Acid sequence window and resubmit. Compare the results from the two searches and determine the potential sites that are different in the insertion and deletion alleles, and therefore altered, by the insertion-deletion polymorphism. Which site is specific to the insertion?

Answer Key

Cross-database Search

1. Geo Profiles
3. SNP = 563 and Gene = 14
4. The search hits aliases, descriptions of interacting proteins, and sequences from other species.

Entrez Gene

7. NOS2A is on chromosome 17
8. 5' LOC441789 and 3' LOC201229

dbSNP

2. How many cSNPs are found in dbSNP for NOS2A? 19 (12-synonymous, 7-nonsynonymous) How many cSNPs are validated? Eight. (828036 is listed twice with two rs numbers)
4. 294
6. 13 SNPs HapMap confirmed
7. Hint: count the number without this data and subtract from total: 244
8. Multiple independent submissions and allele frequency or genotype data.
9. 2, EGP and BCM
11. AD (African Descent) = G = 0.857, C = 0.143; ASIAN = G = 0.818, C = 0.182; YORUBAN = G = 0.875, C = 0.125; CEPH = G = 0.955, C = 0.045

Entrez SNP

3. 334 SNPs
10. 46
13. 15 (if you don't release previous search) if you clear Validation: 2hit-2allele then 215
15. How many total entries are in dbSNP for this gene and submitted by the NIEHS EGP project (our handle is EGP_SNPS)? 252

HapMap Browser:

4. How many HapMap SNPs in this gene? 51
11. rs4462652 , rs2274894, rs3729508, rs4795067, rs2297516, rs944725, rs8068149, rs8072199

GeneSNPs

3. 3' to 5' or right to left.
5. The first position of codon 163 causing an isoleucine to valine substitution.
6. The PDR, G/G

NIEHS SNPs Variation Data:

1. How many non-synonymous cSNPs were discovered in this gene? Four. What is the position of the first synonymous SNP location in our reference sequence? 3816
2. What was the cDNA position of this SNP? 38
3. In which population was it discovered? European
4. Go 'BACK'. In the 'Genotyping' category, click on the 'Visual Genotype' link. An image of all the genotyping data for this gene is displayed. Using the SNP location of the synonymous SNP, determine which individual carries this polymorphism? E119

Genome Variation Server:

No answers to report.

Polyphen

6. Q356R, Q356R and S1040N.
7. P073

PolyDoms

3. HGBASE, SNP500 Cancer, and Sequenome.
4. N1040S
8. 345

ECR Browser and Trafac

4. NBR2, 3' to 5', opposite of BRCA1.
12. NF-1B1, 2, yes.
13. Identifying SNPs in the potential TFBS requires tools beyond these websites, since the alignment of the consensus TFBS is imperfect. With a default 80% sequence alignment to an often degenerate TFBS sequence it is difficult to find the alignment in the original sequence, and once located, it should be determined whether the SNP position is conserved in both the alignment to the consensus TFBS and between the human and mouse sequences. The (-) strand, which means the sequence displayed in the alignment. The (+) strand is the reverse complement of the displayed sequence.
15. GATA-1 at 77, 78 (the site is listed twice by two entries in Transfac)